

# A Dramatic Difference in Global Gene Expression between TCDD-Treated Atlantic Tomcod Larvae from the Resistant Hudson River and a Nearby Sensitive Population

Stuart M. Brown<sup>1,2</sup>, Adriana Heguy<sup>3,4</sup>, Paul Zappile<sup>3,4</sup>, Hao Chen<sup>1</sup>, Aayush Goradia<sup>1</sup>, Yilan Wang<sup>1</sup>, Yuhan Hao<sup>1,2</sup>, Nirmal K. Roy<sup>5</sup>, Kristy Vitale<sup>5</sup>, R. Christopher Chambers<sup>6</sup>, and Isaac Wirgin<sup>5,\*</sup>

<sup>1</sup>Department of Cell Biology, NYU School of Medicine

<sup>2</sup>Center for Health Informatics and Bioinformatics, NYU School of Medicine

<sup>3</sup>Genome Technology Center, NYU School of Medicine

<sup>4</sup>Department of Pathology, NYU School of Medicine

<sup>5</sup>Department of Environmental Medicine, NYU School of Medicine

<sup>6</sup>Howard Marine Sciences Laboratory, Northeast Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, Highlands, New Jersey

\*Corresponding author: E-mail: isaac.wirgin@nyumc.org.

Accepted: August 22, 2017

## Abstract

Atlantic tomcod in the Hudson River Estuary bioaccumulate high hepatic burdens of environmental toxicants. Previously, we demonstrated that Hudson River tomcod developed resistance to TCDD and PCB toxicity probably through strong natural selection during their early life-stages for a variant of the Aryl Hydrocarbon Receptor2 (AHR2). Here, we evaluated the genomic consequences of the resistant genotype by comparing global gene expression in larval tomcod from the Hudson River with expression in larvae from a nearby sensitive population (Shinnecock Bay). We developed an annotated draft tomcod genome to explore the effects of multigenerational exposure to toxicants and a functionally impaired AHR2 on the transcriptome. We used the tomcod genome as a reference in RNA-Seq to compare global gene expression in tomcod larvae from the Hudson River and Shinnecock Bay after experimental exposure of larvae to graded doses of TCDD. We found dramatic differences between offspring from the two populations in the number of genes that were differentially expressed at all doses (0.01, 0.1, and 1 ppb) and even in the vehicle controls. At the two lowest TCDD doses, 250 and 1,141 genes were differentially expressed in Shinnecock Bay larvae compared with 14 and 12, respectively, in Hudson River larvae. At the highest dose (1.0 ppb), 934 genes were differentially expressed in Shinnecock Bay larvae and 173 in Hudson River larvae, but only 28 (16%) of affected genes were shared among both populations. Given the large difference between the two populations in the number and identity of differentially expressed genes, it is likely that the AHR2 pathway interacts directly or indirectly with many genes beyond those known in the AHR2 battery and that other regulatory systems may also respond to TCDD exposure. The effects of chronic multi-generational exposure to environmental toxicants on the genome of Hudson River tomcod are much greater than previously expected.

**Key words:** RNA-Seq, aryl hydrocarbon receptor, transcriptome, draft genome, environmental toxicants, bioaccumulation.

## Introduction

Prior to federal enactment of the U.S. Clean Water Act in 1972, aquatic ecosystems were often the final sink for contaminants released from industrial facilities. Fishes in waterways near urban and industrial locales are often exposed to persistent toxicants that accumulate to extraordinary high levels in sediments, bioaccumulate in populations, and

biomagnify at the apex of aquatic food chains. The Hudson River Estuary in the New York City Metropolitan area has a long history of anthropogenic disturbances from the release of toxic chemicals from both local and distant sources (Wirgin et al. 2006). These include polychlorinated biphenyls (PCBs), polychlorinated dibenzo-*p*-dioxins and furans (PCDD/Fs), polycyclic aromatic hydrocarbons (PAHs), and a variety of toxic

© The Author 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

metals. PCBs have been of particular concern for this ecosystem and this is reflected in the designation of the Hudson River PCBs Superfund site, the nation's largest which encompasses over 300 linear km of river and adjoining floodplain (Thomann and St. John 1979). Sediment levels of PCDD/Fs, notably the most toxic congener, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), are at world record levels in the western Hudson River estuary as a consequence of the operation of an herbicide production facility on the banks of the Passaic River, a tributary of this system. Contamination with TCDD has included downstream waters in Newark Bay and the nearby Hackensack River which has resulted in the designation or proposed designation of these waterways as additional U.S. federal Superfund sites (Quadrini et al. 2015).

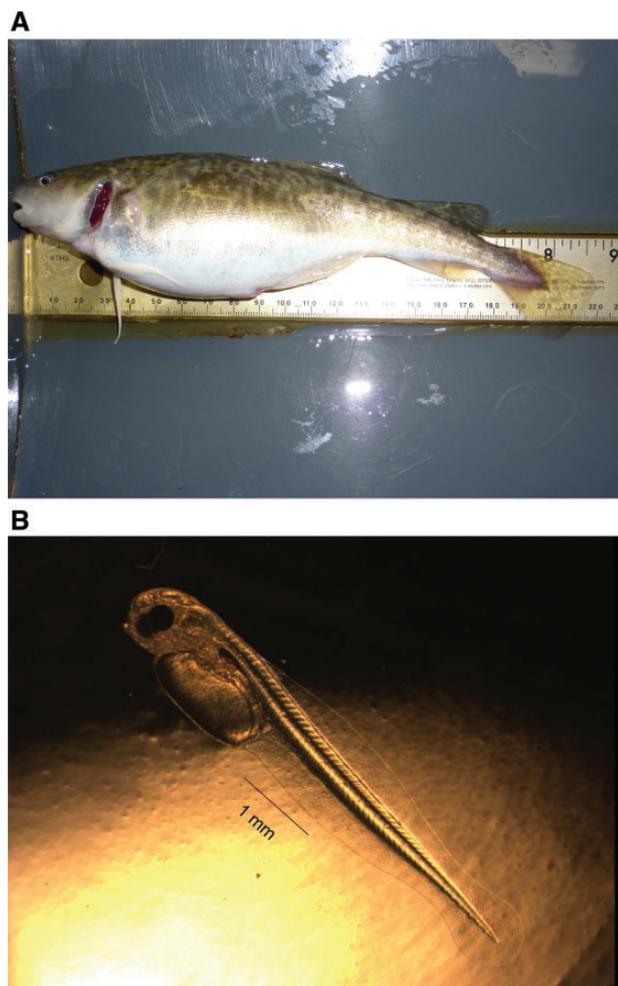
Polychlorinated biphenyls were manufactured primarily for use as insulating fluid in electrical equipment (Breivik et al. 2002) whereas PCDD/Fs are byproducts of a variety of manufacturing and thermal processes including herbicide production, incineration of municipal and medical waste, and pulp and paper production (Rappe 1993). These chemicals are lipophilic, often refractory to metabolism, highly persistent in environments, and biomagnify in food chains (Henny et al. 2009). PCBs and PCDD/Fs exist in many different structural forms (congeners) with levels of environmental persistence that vary and depend largely on the number and molecular location of their chlorine substitutions. Similarly, the toxicities of the congeners vary widely, with those that are coplanar and most closely resembling TCDD being the most toxic (Van den Berg et al. 1998). Known toxicities of coplanar PCBs and PCDD/Fs include tumor promotion, teratogenicity, immunotoxicity, and endocrine disruption.

Toxicities of coplanar PCBs, PCDD/Fs, and some PAHs are mediated by the aryl hydrocarbon receptor (AHR) pathway which, when activated by exogenous ligands such as TCDD and coplanar PCBs, translocates from the cytoplasm to the nucleus where it initiates transcription of a number of xenobiotic (drug)-metabolizing genes in the AHR battery. A variety of endogenous ligands are also known to be AHR agonists (Denison and Nagy 2003; Nguyen and Bradfield 2008). The AHR pathway structure and function is largely conserved among vertebrates. Fishes (Hahn et al. 1997) and birds (Yasui et al. 2007), however, contain two AHRs (AHR1 and AHR2), of which AHR2 is usually more highly expressed, better binds ligands such as TCDD, and therefore is considered more functional in activating gene expression and mediating toxicities (Hahn et al. 1997). Genes in the AHR battery are known to contain multiple dioxin response elements (DREs) in their upstream promoters to which AHR complex binds and initiates their transcription. Although transcriptional activation of genes in the AHR battery is a first step and links directly to some toxicities of these compounds, it is also known that the AHR pathway interacts with several other regulatory pathways resulting in a variety of other phenotypic changes (Tappenden et al. 2013).

The focal species of our study, Atlantic tomcod *Microgadus tomcod*, is anadromous with populations distributed in Atlantic coast estuaries from Labrador to the Hudson River. Tomcod spawn at the freshwater interface of estuaries including the main stem Hudson River and Hackensack River during early winter. They are bottom-dwelling, have lipid-rich livers, and are resident year-round in their natal estuaries—characteristics that render them vulnerable to the bioaccumulation and toxicities of lipophilic contaminants such as PCBs and PCDD/Fs (Wirgin and Chambers 2006). Tomcod are at a critical node in the Hudson River food web because of their unique wintertime spawning making their young-of-the-year an ideal size for consumption by the larger members of the fish fauna in the following summer months (Wirgin and Chambers 2006).

At one time, tomcod from the Hudson River exhibited an unusually high prevalence of liver tumors (hepatocellular carcinomas) and a truncated age structure compared with other tomcod populations from less contaminated rivers (Smith et al. 1979; Dey et al. 1993). More recent studies demonstrated that tomcod from the Hudson River, in contrast to tomcod from other locales, are largely resistant to early life-stage toxicities from exposure to coplanar PCBs and TCDD, but not to PAHs (Wirgin and Chambers 2006). This variation in susceptibility to toxicity was accompanied by a magnitude 100-fold reduction in inducibility of cytochrome P4501A (CYP1A), a well characterized and highly inducible gene in the AHR battery (Wirgin et al. 1992; Courtenay et al. 1999; Yuan et al. 2006a, 2006b). Furthermore, we demonstrated that the resistant phenotype was heritable and caused by deletion of two amino acids just downstream of the ligand binding domain of AHR2 (Wirgin et al. 2011). The AHR2 variant allele was nearly fixed in the Hudson River and Hackensack River (a contaminated tributary to the Hudson River Estuary) populations, but absent in other populations coastwide except for two in closest geographic proximity to the mouth of the Hudson River (Niantic River, Connecticut and Shinnecock Bay, New York) where it occurs at a <5% frequency, and only as heterozygotes (Wirgin et al. 2011).

Because tomcod exhibit a life history that increases the likelihood of contaminant exposure, reside in both clean and contaminated estuaries, has evolved a PCB resistant phenotype, and is an important forage fish of estuarine food webs, we believe that it is an important and potentially informative model with which to investigate the consequences of contaminant exposures and the mechanistic bases of their toxicities. Our objectives in this study were twofold. First, we sought to develop an annotated tomcod genomic DNA sequence with which to evaluate the effects of chemical exposure on global gene expression and the mechanisms whereby it is regulated. Second, we sought to use our newly developed draft tomcod genome to compare levels of global gene expression in larvae from the resistant Hudson River



**FIG. 1.**—Image of a (a) ripe female Atlantic tomcod from the Hudson River used to produce the (b) larval offspring exposed to graded doses of TCDD in this study.

tomcod population treated with graded doses of TCDD compared with similarly treated larvae from the more sensitive nearby Shinnecock Bay population in eastern Long Island, New York. We hypothesized that all genes downstream of the AHR2 pathway would be differentially expressed (DE) in tomcod from Shinnecock Bay at the lowest doses tested, but not in tomcod from the Hudson River, until exposed to sufficiently high doses of  $\sim 100$ -fold higher based on our earlier experiments. This led to two predictions: 1) the number of genes DE at lower doses would be greater in Shinnecock Bay larvae than Hudson River larvae, and 2) the identity of DE genes would differ between Shinnecock Bay larvae and Hudson River larvae because of reduced functionality of the AHR2 pathway in tomcod of Hudson River ancestry. To address these objectives, larvae from both populations were treated with graded doses of TCDD and acetone vehicle control and RNA-Seq was used to identify DE genes.

## Materials and Methods

### Tomcod Collections and Husbandry

Mature adult tomcod (fig. 1a) were collected with unbaited boxtraps from two source populations. The Hudson River tomcod were collected at West Point, New York (river km 85; 41.395063,  $-73.952408$ ) during January, 2012 (fig. 2). Tomcod from Shinnecock Bay ( $\sim 135$  km east of the mouth of the Hudson River along the south shore of Long Island, New York; 40.877924,  $-72.488480$ ) were collected in late December, 2011 (fig. 2). All tomcod were transported to the NOAA Northeast Fisheries Science Center's Howard Marine Sciences Laboratory where ripened gametes were collected, fertilizations were conducted, and embryos were maintained as described previously under NYU IUCAC Protocol 060504-03 (Wirgin and Chambers 2006). Embryos from each population were maintained separately to hatch and beyond for subsequent chemical exposures.

### Chemical Exposures

Groups of 20 tomcod larvae (Shinnecock Bay) and 50 larvae (Hudson River; fig. 1b) that were  $\sim 60$  days post fertilization from each source population were waterborne exposed in 50 ml of artificial sea water (5 parts per thousand salinity) to three doses (0.01, 0.1, and 1.0 parts per billion (ppb) of TCDD each in triplicate beakers (Accustandard, New Haven, CT; Cat#-D4045; purity-99.1%) or acetone vehicle control (also in triplicate beakers) for 7 days. After exposure, pools of 20–25 larvae from each beaker at each dose of TCDD or acetone control were collected in 1.5-ml microcentrifuge tubes, snap frozen, and stored at  $-80^{\circ}$  C until processing.

### RNA Isolation

Total RNAs were isolated from two of the three replicate pools of larvae for each of the three doses of TCDD and the acetone control from each of the two populations using Ultraspec reagent (Biotex, Houston, TX) as described in Yuan et al. (2006a). Because each pool of larvae for each dose of TCDD or acetone control were from different replicate beakers, they represented true biological replicates. RNA concentrations and purities were evaluated using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technol. Inc., Wilmington, DE). The integrity of the RNA was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA).

### DNA Sequencing

Genomic DNA was isolated from dorsal fin clips from two adult Hudson River specimens using a Qiagen QIAamp DNA Mini Kit (Qiagen, Germantown, MD). A whole genome shotgun library was prepared as follows: 500 ng of genomic DNA was sheared to 500 bp fragments using a Covaris E220



**FIG. 2.**—Map depicting the Hudson River and Shinnecock Bay, New York, locales where adult Atlantic tomcod were collected for breeding larval offspring exposed to graded doses of TCDD in this study.

ultrasonicator (Woburn, MA) and the library was prepared using the Kapa Low-Throughput Library Preparation Kit Standard (Kapa Biosystems, Wilmington, MA), without PCR amplification. The library was sequenced as 101 base paired-end reads on an Illumina (San Diego, CA) HiSeq 2500 (v4 chemistry) instrument with a yield of 210 million paired-end reads (42 Gb). Initial quality control inspection with FastQC (Andrews 2010) showed 90% of bases  $>Q30$  and a 45% GC content. A sequence duplication level of 25% was found, which is higher than expected for the estimated  $40\times$  genome coverage but may indicate some repetitive DNA content in the genome. DNA reads had primer/adaptor sequences removed and were then quality trimmed from both ends with Trimmomatic (Bolger et al. 2014) until the final 4-base window had an average  $Q30 \geq 15$ .

Synthetic long-read libraries were built using the Illumina TruSeq Synthetic Long Read Library prep kit. High-molecular-weight genomic DNA ( $>40$  Kb) was sheared into 10-Kb size fragments using a Covaris g-Tube spun at  $4,200 \times g$ . Illumina adapters were ligated to the 10 Kb fragments, and the ligated fragments size selected using an E-Gel and a Qiagen (Valencia, CA) QIAquick gel extraction kit. Following quantification using QPCR, Qubit, and the high sensitivity Bioanalyzer, the 10 Kb fragments were diluted and distributed across a 384-well plate, and clonally amplified by long-range PCR. The amplified fragments were fragmented to about 350 bp using a “tagmentation” procedure via a transposase. Unique indices (barcodes) were assigned to each well by indexing PCR and the fragments from all wells were pooled

and concentrated using a Zymo (Zymo Research, Irvine, CA) column. The pool was sequenced in one lane with HiSeq 2500 (v4 chemistry) as  $2 \times 100$  reads. The sequences were uploaded to BaseSpace (Illumina) and processed using the TruSeq Long-Read Assembly App (Illumina), which preprocesses the short-reads to correct for PCR and sequencing errors, and assembles them into contigs using the string Graph Assembler (SGA), followed by assembly of contigs into scaffolds. The end result was a set of synthetic long reads representing 10 Kb genomic fragments.

De novo assembly of the standard Kapa 101 base paired-end read library, supplemented with Illumina synthetic long reads was conducted with SOAPdenovo2 (v 2.0.4; Luo et al. 2012). The draft genome contained a total of 422 Mb in contigs and 679 Mb in scaffolds (including predicted gaps). The scaffold N50 was 20,585 bp, with a total of 174,603 contigs. Genes were ab initio predicted in the draft contigs with GLIMMER-HMM (v. 3.0.4), yielding a total of 17,648 genes. Predicted genes were translated and amino acid sequences were compared with BLAST to UniProt proteins and to all zebrafish RefSeq proteins. Predicted proteins were also annotated by hmmscan (HMMER3 profile hidden markov model search (Eddy 2011)) to the complete Pfam library of protein motifs (Finn et al. 2016).

### RNA Sequencing

We constructed a transcriptome for the tomcod using a combination of predicted gene coding sequences predicted from the draft genome with GLIMMER-HMM (Majoros et al. 2004)

and de novo transcript assemblies built from RNA-seq libraries collected at different developmental stages. To build a broadly representative transcriptome, a developmental time course of RNA samples was isolated from livers of two environmentally exposed adult Hudson River tomcod, livers of two lab-reared juvenile Hudson River tomcod that were treated with 1 ppm of benzo[a]pyrene (B[a]P), two pools of five larvae each of Hudson River ancestry that were exposed to 1 ppm B[a]P, and two pools each of five 14-days-old embryos of Hudson River ancestry that were waterborne exposed to 1 ppm B[a]P. RNAs were isolated, quantified, and evaluated as described above. RNA-Seq libraries were prepared using the Illumina TruSeq mRNA v2 kit or the TruSeq stranded mRNA kit, with 12 cycles of PCR amplification, starting from 500 ng of total RNA. Barcoded libraries were sequenced as 101 base paired-end reads on the Illumina HiSeq 2500 (v4 chemistry).

RNA-seq libraries for each developmental stage were assembled separately with rnaSPAdes (Bankevich et al 2012) and with SOAPdenovo-Trans (Xie et al. 2014). The genomic gene predictions and the RNA transcript assemblies were combined into a single set of transcripts with the EvidentialGene pipeline (Gilbert 2013). The nonredundant set of transcripts was then screened by alignment with Diamond (Buchfink et al. 2015) against the complete UniRef50 database of proteins (Suzek et al. 2015), and only sequences with a match with e-value smaller than  $e < 10^{-3}$  were retained. The final set of tomcod transcripts contained 25,522 coding sequences. Additional annotation was provided by matching these transcripts to the complete set of ENSEMBL zebrafish proteins (*Danio rerio*) with BLAST at a threshold of  $e < 10^{-3}$ .

Gene expression in the TCDD treated and control samples from the two populations was quantified by Salmon using the quasi-mapping-based mode (Patro et al. 2017) using the tomcod transcriptome as the reference. Hudson River versus Shinnecock Bay larval pools were compared at each dose level (acetone negative control, 0.01, 0.1, and 1.0 ppb of TCDD), and each TCDD treated pool was compared with the control from each of the two populations. Gene counts were compared for differential expression between sample groups using edgeR (Robinson et al. 2010).

Functional analysis of the DE genes was conducted as follows. BLAST was used to map all of the tomcod genes to zebrafish (*Danio rerio*) gene IDs in the ENSEMBL database (Aken et al. 2016). Lists of significant DE genes for each treatment vs. control were analyzed by DAVID (Dennis et al. 2016) (<https://david.ncifcrf.gov/home.jsp>) for enriched Gene Ontology functions. DAVID combined analysis of function and reported clusters of function terms with scores for enrichment and FDR corrected *P*-values set at  $< 0.05$ .

### Q-RT-PCR Validation of DE Genes

Total RNA was extracted from two pools of tomcod larvae each from the three replicated treatment groups of both

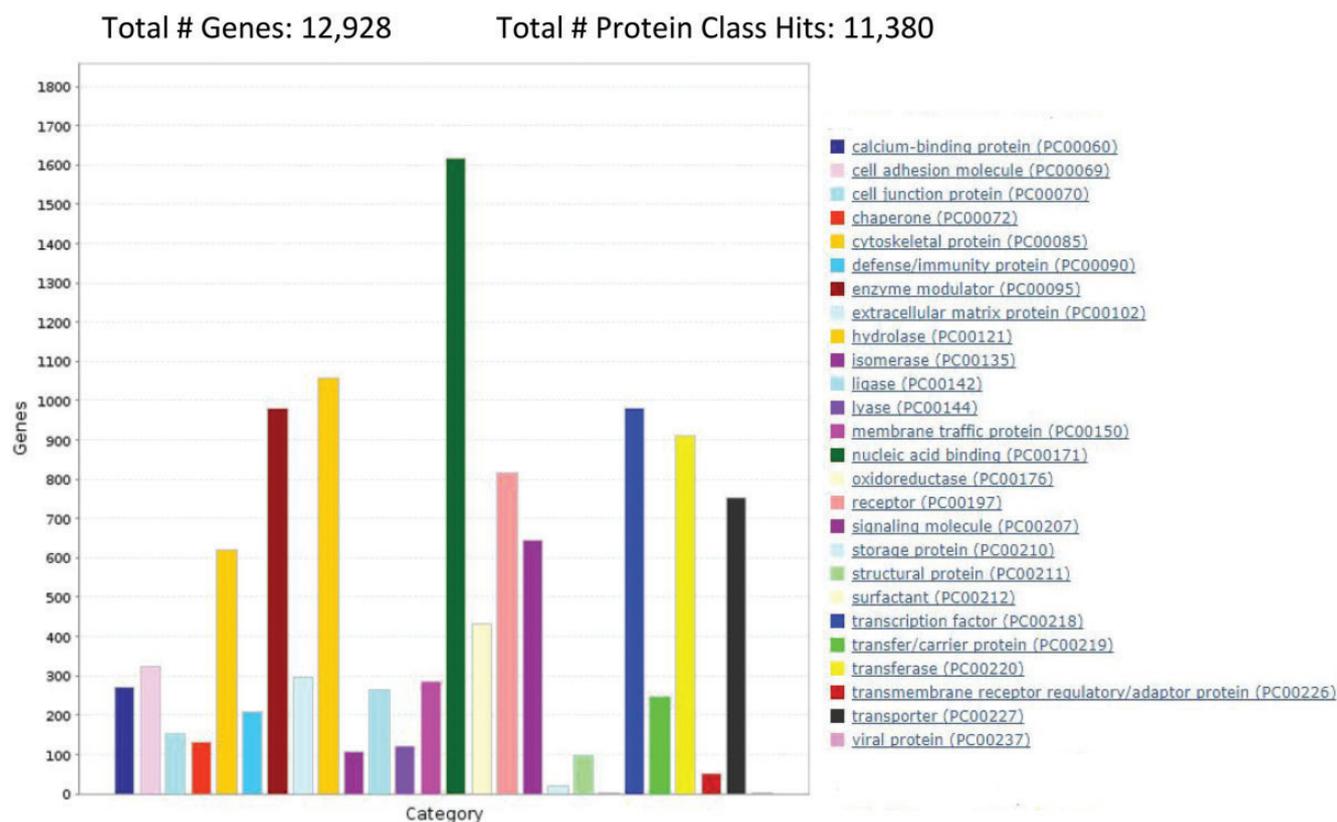
Hudson River and Shinnecock Bay larvae that were waterborne exposed to three doses each of TCDD (0.01, 0.1, and 1.0 ppb) and acetone control using TRIzol Reagent (Life Technologies) according to the manufacturer's protocol. Thus, RNA was analyzed for gene expression from 4 to 6 pools of tomcod larvae for each of three doses of TCDD and acetone control from each population. Total RNA concentrations for samples were measured in a NanoDrop2 spectrophotometer (Thermo Fisher Scientific) and used as a template for first strand cDNA synthesis using M-MLV Reverse Transcriptase (Promega, Madison, WI) as described in Roy et al. (2011).

Real-time PCR primers were designed based on our derived tomcod RNA-Seq sequences (supplementary table 1, Supplementary Material online). RT-PCR reactions were done as described by Roy et al. (2011) with SYBR-green PCR Master Mix (Fisher Scientific, Pittsburgh, PA) in a QuantStudio 6 Real-Time PCR System instrument (Thermo Scientific) under the following cycling conditions: 10 min at 90 °C, then 40 cycles at 95 °C for 15 s and 60 °C for 60 s in 384-well plates. Each individual sample was analyzed in triplicate.

The comparative Ct method ( $\Delta\Delta$ Ct Method) was used to quantify gene expression by normalizing Ct values against a tomcod  $\beta$ -actin housekeeping gene and comparing cycles for chemical treatment groups to the acetone control group for each population. The six values for fold changes for each dose for each population were averaged and this mean was used to represent the fold change in expression within and between populations ( $\pm$  SD). Statistical comparisons between the Hudson River and Shinnecock Bay populations as a whole, at each dose within and between populations compared with their own acetone controls, was made with ANOVA in Excel.

## Results

Our initial approach to study global gene expression in Atlantic tomcod was to build a draft genome assembly, which could then be used to align RNA-Seq reads and annotate genes. Illumina paired-end 100 bp reads were collected from genomic DNA from adult Hudson River tomcod livers and assembled into contigs and scaffolds with the SOAPdenovo assembler. Additional synthetic long read contigs were included in the assembly process. The draft tomcod genome contained a total of 422 Mb in contigs, and 679 Mb in scaffolds (including predicted gaps). The scaffold N50 was 20,585 bp, with a total of 174,603 contigs. The closely related Atlantic cod *Gadus morhua* has a draft genome assembly of 608 Mb (total assembly length of 824 Mb with predicted gaps), but it does not have a complete set of annotated or predicted proteins. Genome size varies widely among teleost fish, from 391 Mb for fugu *Takifugu rubripes* to 1,713 Mb for common carp *Cyprinus carpio*. The well-studied model organism, zebrafish, has a genome size of 1,392 Mb which encodes



**Fig. 3.**—Bar chart of predicted protein functions of 22,522 predicted Atlantic tomcod coding sequences identified in this study.

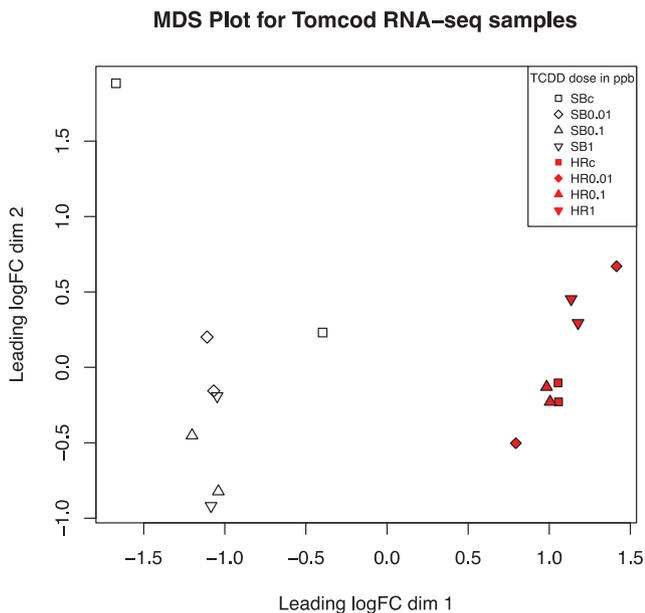
24,297 RefSeq nonredundant proteins. Among fish, genome size does not seem to correlate with the number of encoded proteins, nor any other phenotypic trait.

To further annotate RNA transcripts, we performed RNA-Seq and de novo transcriptome assembly. RNA samples were isolated from tomcod embryos, larvae, juvenile livers, and adult livers. These samples (one pooled sample from each developmental stage with no replicates) were sequenced with the standard Illumina RNA-seq protocol (101 bp paired-end). RNA-seq libraries for each developmental stage were assembled with rnaSPAdes (Bankevich et al 2012). The genomic gene predictions and the RNA transcript assemblies were combined into a single set of transcripts with the EvidentialGene pipeline (Gilbert 2013) for a total of 25,522 coding sequences. The list of genes was filtered to only include those with a significant ( $1e^{-3}$ ) BLAST match to UniProt. The complete set of predicted proteins were further annotated by hmmscan with all Pfam protein family motifs. Figure 3 shows a bar chart of the most common Pfam functions grouped by Gene Ontology categories (26 Protein Classes [11,380 hits]) and supplementary table 2, Supplementary Material online contains a complete list of the 25,522 tomcod predicted proteins and their annotations. Completeness of the draft genome was assessed by BLAST with the CORE highly conserved eukaryotic genes (Korf

2015). The draft genome contains high quality blast matches (e-value of  $1e^{-20}$  or better over >50% of the query length) for 99% of the eukaryotic CORE genes (Parra et al. 2009). We found that 92% of known zebrafish proteins (from GenBank) had a high-quality BLAST match with the tomcod transcripts, suggesting that the tomcod transcriptome is nearly complete.

The tomcod transcriptome was used for gene expression analysis using RNA-Seq of the response of larval tomcod of Hudson River and Shinnecock Bay ancestry to waterborne exposures for 7 days to three graded doses of TCDD or acetone solvent control. RNA sequencing was conducted by the standard Illumina mRNA library protocol. RNA-Seq reads were aligned to the tomcod transcriptome and per-gene expression was quantified. Differential expression was calculated between acetone control treated Hudson River and Shinnecock Bay larvae, and separately for each TCDD dose compared with the control from the same population in duplicate (all data summarized in supplementary table 3, Supplementary Material online).

Levels of gene expression were significantly different between Hudson River and Shinnecock Bay larvae, regardless of the level of TCDD treatment and in acetone treated controls. Moreover, largely distinct sets of genes were DE between specimens from the two locations at all doses of TCDD and even acetone solvent controls. Furthermore, the modest number of genes that were DE in Hudson River larvae at the



**Fig. 4.**—MDS plot of global gene expression detected by RNA-Seq in Hudson River (HR) and Shinnecock Bay (SB) larvae exposed for 7 days in duplicate treatment groups to three graded doses of TCDD (0.01, 0.1, and 1.0 ppb) and acetone control (c). The largest principle component of variation (x axis) separates the HR from SB samples. All three doses of TCDD elicited differential expression of a large number of genes compared with acetone control in SB larvae whereas the two lowest doses of TCDD elicited only a small number of differentially expressed genes compared with acetone control in the HR treatment groups.

highest dose of TCDD were for the most part distinct from those DE in Shinnecock Bay at low (0.01 ppb) or high (1.0 ppb) TCDD dose. The MDS plot of all samples in this study (fig. 4) showed the largest source of variance in the RNA-Seq data was the location of the source larval populations (i.e., Hudson River vs. Shinnecock Bay).

The lists of DE genes for treatment by TCDD versus acetone controls for each population are shown sorted in supplementary material (Shinnecock Bay [supplementary tables 4–6]; Hudson River [supplementary tables 7–9]), Supplementary Material online, by *q*-value and then by log-fold change. Our most striking result was that Shinnecock Bay larvae showed a much stronger response to TCDD exposure at lower doses than those of Hudson River origin, starting at the lowest dose tested. Shinnecock Bay larvae exhibited 250 DE genes at 0.01 ppb TCDD (138 up; 112 down), 1,141 DE genes at 0.1 ppb TCDD (355 up; 786 down), and 934 DE genes at 1.0 ppb (219 up; 715 down; fig. 5a and 6; supplementary tables 4–6, Supplementary Material online). In contrast, in Hudson River larvae only 14 DE genes were found at the 0.01 ppb dose (13 up; 1 down) and 12 DE genes at the 0.1 ppb dose (12 up; 0 down; supplementary tables 7 and 8, Supplementary Material online). There were 173 DE genes in the Hudson River larvae at the highest dose of TCDD used,

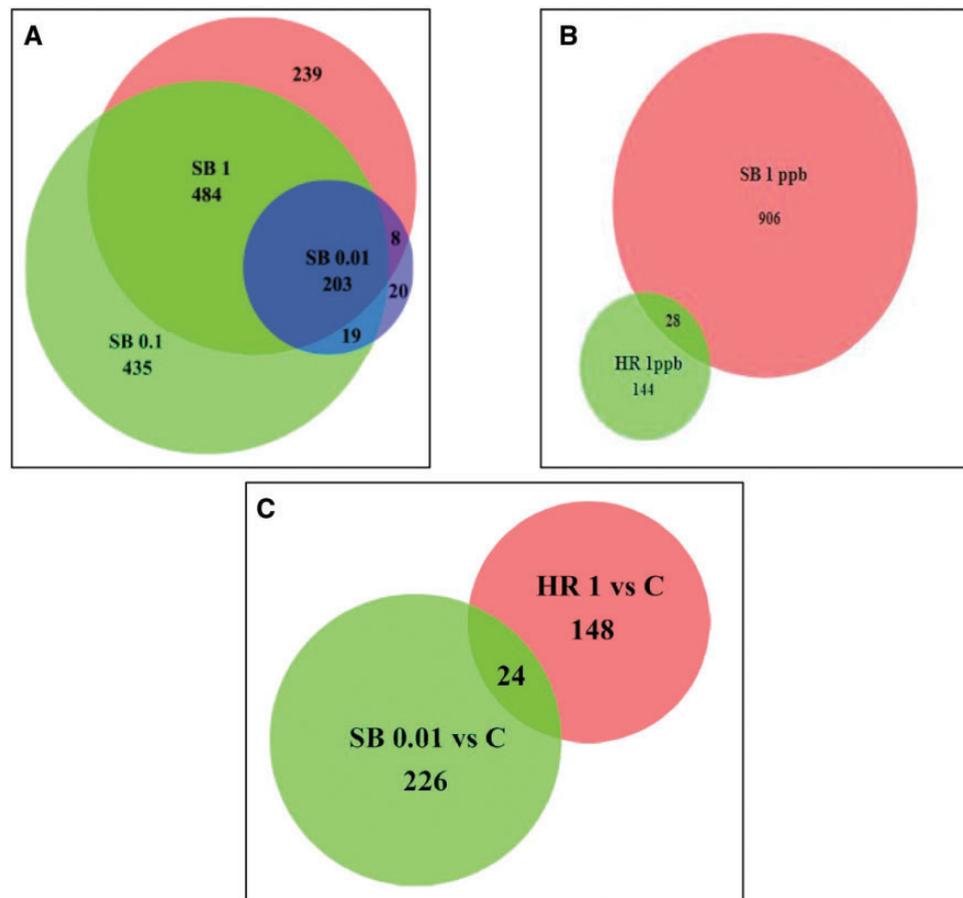
1.0 ppb (88 up; 85 down; fig. 7; supplementary table 9, Supplementary Material online).

In Shinnecock Bay larvae, there was remarkable consistency across TCDD doses with 687 of 934 DE genes (74%) at the highest dose (1.0 ppb) also found DE at the intermediate dose (0.1 ppb), and 222 of the 250 (88%) DE genes at the lowest TCDD dose (0.01 ppb) also found to be DE at the intermediate dose. This is visualized as a largely overlapping Venn diagram of DE genes in the three TCDD treatments of Shinnecock Bay larvae (fig. 5a). In the Shinnecock Bay larvae, the top 25 upregulated DE genes at 0.01, 0.1, and 1.0 ppb TCDD were almost identical (supplementary tables 4–6, Supplementary Material online). We observed in SB larvae that multiple cytochrome P450 genes and heat shock proteins, were among the most upregulated genes at all doses of TCDD and multidrug resistance genes were significantly downregulated. For example, 7 of the 17 most upregulated genes in SB larvae at 0.1 ppb TCDD were cytochrome P450 genes and the AHR repressor.

In contrast, there was little overlap between the TCDD-induced genes in Hudson River and Shinnecock Bay larvae. Only 28 of the 173 (16.2%) DE genes in Hudson River larvae at the highest TCDD dose (1.0 ppb) were also found DE in Shinnecock Bay larvae ( $n = 934$ ) at that dose (fig. 5b). Similarly, there was very little overlap when comparing the high dose (1.0 ppb) exposed Hudson River larvae to the low dose (0.01) exposed Shinnecock Bay (fig. 5c). Only 24 DE genes were shared between the two samples (13.9% of Hudson River DE genes).

Results from RNA-Seq analysis were validated by Q-RT-PCR from an expanded number of specimens from the two populations ( $n = 6$  pools of larvae each) for each of the three doses of TCDD at nine loci for which DE genes were identified by RNA-Seq (table 1). Significant differential expression was observed between Hudson River and Shinnecock Bay larvae at all three TCDD doses at four loci (CYP1A1, CYP1B1, CYP1C1, and CYP1C2) and a subset of TCDD exposure doses at the five remaining loci (RALDH2, Troponin T2E [cardiac], Ca Activated K<sup>+</sup> Channel Subunit Alpha-1 isoform, X25, ATP2A1). Similarly, differential expression compared with the vehicle acetone control group was observed within both the Hudson River and Shinnecock Bay larvae at four of the loci (CYP1A1, CYP1B1, CYP1C1, and CYP1C2), but only within the Shinnecock Bay larvae for RALDH2 and the Hudson River larvae for Troponin T2E (cardiac). The magnitude of gene induction was much higher in the Shinnecock Bay larvae compared with the Hudson River larvae at CYP1A1, CYP1B1, CYP1C1, and CYP1C2.

Functional analysis was conducted using DAVID separately for DE genes in Shinnecock Bay larvae only and Hudson River larvae only treated with each of the three doses of TCDD. We found 21 significantly DE biological processes in Shinnecock Bay specimens treated with the lowest dose of TCDD (0.01 ppb; supplementary fig. 1, Supplementary Material



**FIG. 5.**—Venn diagrams of the number of significantly differentially expressed genes. (a) Between acetone control and low (0.01 ppb), medium (0.1 ppb), and high (1.0 ppb) TCDD doses on duplicate treatment groups of Shinnecock Bay (SB) Atlantic tomcod larvae treated for 7 days. In total, there were 250, 1,141, and 934 significantly differentially expressed genes at the 0.01, 0.1, and 1.0 ppb doses, respectively. Most of the same genes overlap between the different exposure groups. (b) In Hudson River (HR) and SB tomcod larvae exposed for 7 days to the highest dose of TCDD used (1.0 ppb). In addition, to the larger number of significantly differentially expressed genes in larvae of SB ( $n = 934$ ) than those of HR ancestry ( $n = 173$ ) there was little overlap ( $n = 28$ ) in the identity of genes that were differentially expressed in both populations. (c) In HR tomcod larvae treated for 7 days with the highest dose of TCDD (1.0 ppb) and SB larvae treated with the lowest TCDD dose (0.01 ppb). In addition to the larger number of significantly differentially expressed genes in SB larvae ( $n = 250$ ) than those of HR ancestry ( $n = 172$ ), there was little overlap in the identity of genes that were differentially expressed in both populations ( $n = 24$ ).

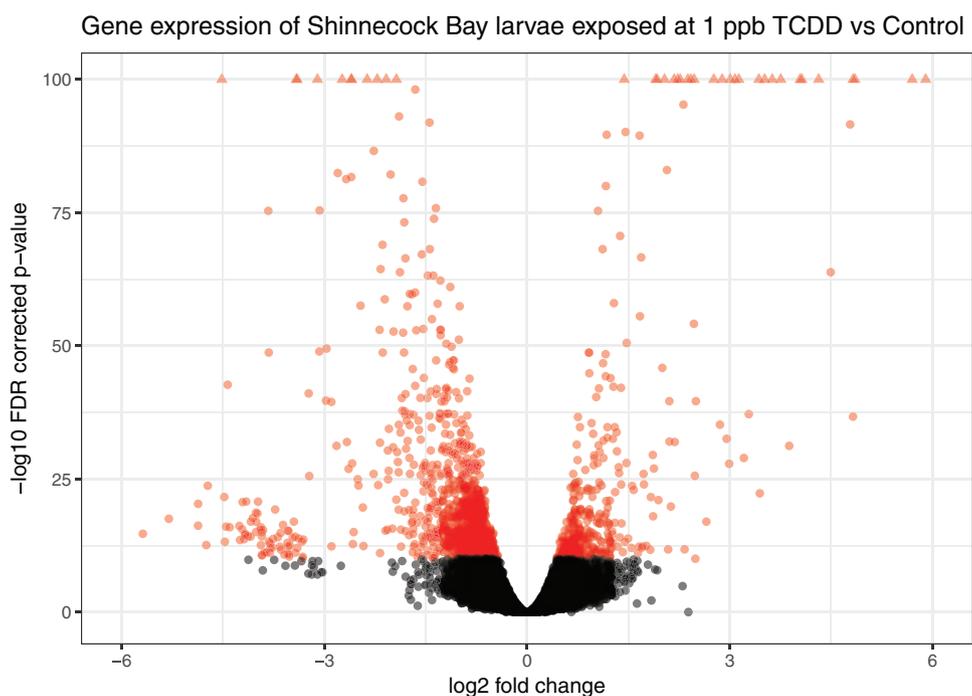
online) and only seven DE biological processes in Hudson River specimens at the highest doses of TCDD (1.0 ppb; supplementary fig. 2, Supplementary Material online). The identity of the functional groups in the DE gene lists differed greatly between the two populations exposed to TCDD and in the negative controls (supplementary figs. 1 and 2, Supplementary Material online). As expected, the functional processes with the highest enrichment scores in Shinnecock Bay larvae included responses to chemical stimuli including xenobiotics, organic substances metabolism, biosynthetic processes, and surprisingly ribosomal related genes (ribosome subunit synthesis, small RNAs, and translation).

A large number of genes were DE between negative control (acetone treated) larvae from the two populations. In total, we observed 1,159 genes that were DE between the negative control larvae from the two populations (575 up

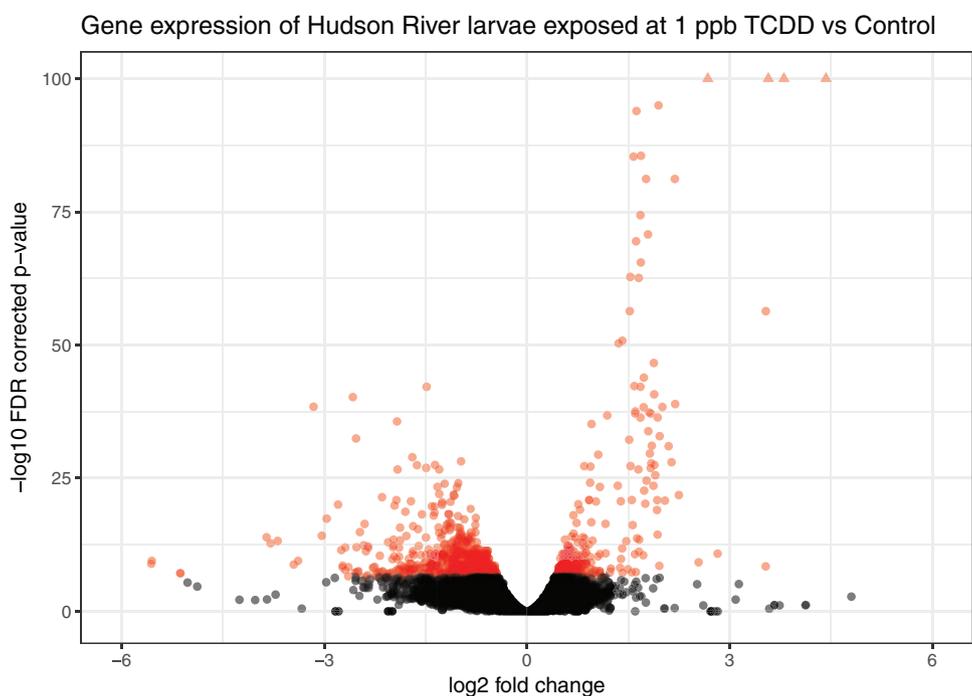
and 584 down comparing Hudson River to Shinnecock Bay larvae (supplementary table 10, Supplementary Material online). This result should be considered in light of the small number of DE genes in Hudson River larvae that were treated with either 0.01 or 0.1 ppb TCDD. DAVID analysis of the functional pathways that differed between untreated larvae from the two populations included enrichment of proteases, lipases, esterases, carbohydrate metabolism, and myosin and filament structural proteins, but did not include the drug and xenobiotic metabolism proteins that were so strongly enriched in the TCDD-treated Shinnecock Bay larvae.

## Discussion

Identification of globally differentially regulated genes and levels of their expression can provide sensitive, mechanistic,



**FIG. 6.**—Volcano plot of Shinnecock Bay larvae treated with the highest dose (1 ppb) of TCDD compared with their respective acetone-vehicle negative controls. Significantly differentially expressed genes with FDR corrected  $P$  values  $< 0.05$  are shaded in red. Outliers with  $-10 \cdot \log(\text{FDR})$  values  $> 100$  are compressed at the top of the plot with triangle symbols.



**FIG. 7.**—Volcano plot of Hudson River larvae treated with the highest dose (1 ppb) of TCDD compared with their respective acetone-vehicle negative controls. Significantly differentially expressed genes with FDR corrected  $P$  values  $< 0.05$  are shaded in red. Outliers with  $-10 \cdot \log(\text{FDR})$  values  $> 100$  are compressed at the top of the plot with triangle symbols.

**Table 1**

Fold Induction Compared with Acetone Control Determined by Q-RT-PCR of Genes Identified as Differentially Expressed by RNA-Seq Analysis in Hudson River and Shinnecock Bay Atlantic Tomcod Larvae That were Exposed for 7 Days to Three Graded Doses of TCDD (0.01, 0.1, and 1.0 ppb) and Acetone Control

Gene (Dose)	Population	
	Hudson River	Shinnecock Bay
CYP1A1 <sup>a,b,c</sup>		
0.01 9.90 ± 4.8	39.57 ± 4.1	
0.1	43.9 ± 4.2	52.02 ± 0.75
1.0	55.9 ± 12.1	90.99 ± 15.5
CYP1C1 <sup>a,b,c</sup>		
0.01	1.32 ± 0.128	11.44 ± 1.16
0.1	1.97 ± 0.214	20.78 ± 1.02
1.0	2.12 ± 0.642	21/67 ± 4.79
CYP1B1 <sup>a,b,c</sup>		
0.01	0.83 ± 0.290	21.81 ± 5.98
0.1	1.54 ± 0.291	56.57 ± 8.56
1.0	1.81 ± 0.050	36.40 ± 6.94
CYP1C2 <sup>a,b,c</sup>		
0.01	0.71 ± 0.145	30.17 ± 4.84
0.1	2.33 ± 0.639	51.07 ± 10.1
1.0	3.81 ± 1.750	54.59 ± 10.5
RALDH2 <sup>a,d,e</sup>		
0.01	1.51 ± 0.147	0.752 ± 0.162
0.1	1.46 ± 0.180	0.678 ± 0.160
1.0	0.92 ± 0.170	0.644 ± 0.050
Troponin T2E (cardiac) <sup>b,e</sup>		
0.01	0.819 ± 0.153	1.00 ± 0.123
0.1	0.927 ± 0.120	1.02 ± 0.109
1.0	0.835 ± 0.159	1.08 ± 0.255
Ca Activated K <sup>+</sup> Channel Subunit Alpha-1 Isoform X25 <sup>c,f</sup>		
0.01	0.543 ± 0.283	1.27 ± 0.320
0.1	0.654 ± 0.050	1.25 ± 0.110
1.0	1.320 ± 0.520	1.19 ± 0.322
ATP2A1 <sup>b,e</sup>		
0.01	0.60 ± 0.358	0.847 ± 0.349
0.1	0.75 ± 0.260	1.586 ± 0.662
1.0	2.00 ± 0.557	1.67 ± 0.82

<sup>a</sup>Significant differential gene expression in all TCDD-treated groups versus acetone control in Shinnecock Bay larvae.

<sup>b</sup>Significant differential gene expression in all TCDD-treated groups versus acetone control in Hudson River larvae.

<sup>c</sup>Significant difference in gene expression between the Shinnecock Bay and Hudson River larvae at all individual doses of TCDD.

<sup>d</sup>Significant differential gene expression in 0.01 ppb TCDD-treated group versus acetone control within Hudson River larvae.

<sup>e</sup>Significant difference in gene expression between the Shinnecock Bay and Hudson River larvae at the 0.01 and 0.1 ppb doses of TCDD.

<sup>f</sup>Significant differential gene expression in 0.01 and 0.1 ppb TCDD-treated group versus acetone control within Hudson River larvae.

and genome-wide information on the early toxic effects of anthropogenically induced environmental stressors. This may be especially valuable when considering sentinel species in compromised environments such as the Hudson River. With the increasing availability of next-generation sequencing,

nontraditional, but imperiled species, such as Atlantic tomcod can be the direct focus of genomic effects studies on impacted natural environments rather than relying on extrapolations from traditional model finfishes such as zebrafish and Japanese medaka *Oryzias latipes*.

Previous efforts to comprehend the full suite of toxicities induced in tomcod by these xenobiotics and their mechanistic bases was limited by the absence of genomic sequence data. The draft annotated sequence generated in this study allows for a global interrogation of the complete suite of genes that individually, or in concert, were DE by exposure to Hudson River-borne toxicants and costressors such as a warming environment (Seekell and Pace 2011). This information provides a greater understanding of the differential effects of exposure to TCDD, PCBs, and complex environmental mixtures, on the genomes of tomcod from two nearby populations with distinctly different exposure histories. In addition, our draft genome provides a nearly complete resource set of proteins, transcripts, and protein coding genomic regions for further candidate gene and genome-wide investigations of impacted and nonimpacted tomcod populations.

This study is among the first to use RNA-Seq to compare global gene expression in offspring of wild fish from a population subjected to chronic exposure for many generations and which has adapted to these toxicants (Wirgin and Chambers 2006; Wirgin et al. 2011). Our major finding, which ran counter to expectation of limited interpopulation differences, is that Hudson River and Shinnecock Bay tomcod larvae exhibit strikingly different basal and TCDD induced patterns of gene expression. The number of toxicant induced DE genes was far greater and at a 100-fold lower dose of TCDD in Shinnecock Bay larvae compared with Hudson River larvae. The Shinnecock Bay larvae had a strong response at the lowest dose of TCDD (0.01 ppb) that affected a large number of genes and metabolic pathways. The response hit a plateau at the intermediate dose (0.1 ppb) with a somewhat lower response at the highest dose of TCDD (1.0 ppb). In contrast, the Hudson River larvae appeared to tolerate low and moderate TCDD doses with very few changes in gene expression, and only began to respond at the highest treatment level (1.0 ppb TCDD). For example, at the lowest dose of TCDD tested (0.01 ppb), 250 genes were DE in Shinnecock Bay larvae and only 14 in the Hudson River group. At the medium dose (0.1 ppb), 1,141 genes were significantly DE in Shinnecock Bay larvae and 12 in the Hudson River group. It was not until the highest dose of TCDD (1.0 ppb) that a moderate number of genes were significantly DE in the Hudson River larvae (173 genes), but still far fewer than the 934 DE genes in Shinnecock Bay larvae.

Surprisingly, the identity of DE genes also differed between the populations with only 28 (3.0%) of 934 DE genes in Shinnecock Bay larvae at the highest TCDD dose shared with Hudson River larvae at the same dose. Even in the acetone vehicle controls, there were 1,159 genes that were

significantly DE between Hudson River and Shinnecock Bay larvae. These population differences suggest that target genes of these xenobiotics were less sensitive to expression changes in the Hudson River larvae and the responsive molecular pathways differed between the populations.

It is often hypothesized that development of resistance, such as that exhibited by Hudson River tomcod, is associated with evolutionary costs that may include among others reduced life expectancy, decreased fecundity, and increased sensitivity to a multitude of other environmental stressors (Whitehead et al. 2017). For example, chemically resistant Atlantic killifish from the creosote-contaminated Elizabeth River, Virginia, exhibited decreased tolerance to low levels of dissolved oxygen and phototoxicity of PAHs (Meyer and DiGiulio 2003), but increased vulnerability to liver and pancreatic cancers (Wills et al. 2010). It is known that the AHR binds a number of endogenous ligands and plays a critical regulatory role in normal vertebrate development in processes such as cell proliferation and vascular, reproductive, and immune system development (Stevens et al. 2009; Safe et al. 2013). Therefore, it is interesting to note the large number of genes that were DE between control larvae from the resistant Hudson River and sensitive Shinnecock Bay populations. It is tempting to speculate that differential expression of a number of these genes may be a developmental “cost” associated with the acquisition of the resistant phenotype in the Hudson River population.

One caveat to our interpretation of the magnitude of the interpopulation differences in gene expression was our use of whole larvae as targets for analysis. Levels of gene expression can differ among tissues of TCDD, PCB, or PAH treated early life-stages of fishes (Carney et al. 2006; Bugiak and Weber 2009; Roy et al. 2011). This intertissue variation may be due to variation in the distribution of contaminants among tissue types or differential sensitivities among them perhaps resulting from differences in levels of expression of components of the AHR pathway. Because, cardiovascular dysfunction is known to precede most developmental toxicities in PAH (Bozinovic et al. 2013) and TCDD (Antkiewicz et al. 2005) treated fishes, our analyses may have been more informative had we used heart tissue alone rather than whole larvae for quantifying gene expression. In future experiments, we plan to compare expression of a subset of DE genes that we identified in this study exclusively in the hearts of TCDD exposed tomcod larvae.

Are the nominal concentrations of TCDD used in our study environmentally relevant? Previously, we quantified hepatic burdens of PCDD/Fs and coplanar PCBs on a congener-specific basis in adult and juvenile tomcod that were environmentally exposed to contaminants in the Hudson River estuary (Courtenay et al. 1999; Fernandez et al. 2004). We reported total hepatic TCDD TEQs (from PCDD/Fs and coplanar PCBs) in juvenile tomcod from Newark Bay and the Hackensack River that exceeded 1.0 ppb. Thus, the

concentrations of TCDD that initially evoked significant differences in gene expression between larvae from the two populations in our current study (0.01 and 0.1 ppb TCDD) were up to two orders of magnitude below that we reported in livers of juvenile tomcod from the Hudson River estuary.

Our results differ from those of several earlier studies that compared global gene expression between populations of Atlantic killifish *Fundulus heteroclitus* that are resistant to TCDD, PCBs, and PAHs in several U.S. Atlantic coast estuaries (reviewed in Wirgin and Waldman 2004). For example, Bozinovic et al. (2013) used microarrays to compare expression levels between  $\beta$ -naphthoflavone-treated resistant killifish embryos from the PAH contaminated Elizabeth River, Virginia to those from a nearby sensitive population (Kings Creek, Virginia). Of the 6,754 genes screened, only 52 (0.76%) were DE because of  $\beta$ -naphthoflavone treatment alone and hierarchical clustering revealed that similar genes were DE in the two populations. Additionally, and also in contrast to our results, only 26 genes (0.38%) exhibited significant differences in expression between  $\beta$ -naphthoflavone treated embryos from the two populations. Oleksiak et al. (2011) used the same microarray platform to compare gene expression in killifish embryos from the PCB-resistant New Bedford Harbor, Massachusetts, population to those from the sensitive Scorton Creek, Massachusetts population that were treated with PCB126 and sacrificed 5, 10, and 15 days postfertilization. They found that 0.38%, 0.40%, and 1.0% of genes were DE between populations at these three time points, respectively. Therefore, the number of genes that were DE between the resistant killifish populations from highly impacted locales described above and their paired sensitive populations were far less than we report here for Hudson River and Shinnecock Bay tomcod larvae.

Results of a recent genomic sequence and transcriptome study of F<sub>2</sub> embryonic killifish offspring from four paired populations, one resistant and one sensitive in each pair, are more closely aligned to ours (Reid et al. 2016). Using RNA-Seq, they found that a greater number of genes (70) were significantly DE between the embryos from the paired resistant and sensitive populations that were treated for 10 days with 0.2 ppb PCB126. They also found that the greater number of upregulated genes in the sensitive compared with resistant populations were enriched for those directly activated by the AHR pathway. In our study, AHR pathway activated genes including cytochrome P450s and Phase II detoxification genes were also among the most highly DE genes in Shinnecock Bay, but not Hudson River larvae. However, many genes not known to be activated by AHR were also DE between the two tomcod populations. This suggests that the phenotypic ramifications of the resistance phenotype in Hudson River tomcod are more far-reaching than previously imagined.

The greater number of DE genes between resistant and sensitive tomcod populations compared with killifish is noteworthy. Some of the interspecific differences may have

resulted from the different chemicals used (PAH vs. PCB126 vs. TCDD), doses, durations of exposures, and different early life-stages tested (embryos vs. larvae). For example, because TCDD is the prototypical and strongest binding AHR2 agonist, it may induce higher levels of gene expression in sensitive populations than the other toxicants and therefore accentuate differences between sensitive and resistant populations. The differences in methods used to measure global expression—microarrays in some of the killifish studies and RNA-Seq here—may have contributed to the reported differences. Compared with microarrays, RNA-Seq is a more sensitive and unbiased method capable of detecting transcripts at low abundance and has a broader dynamic range of transcript detection (Zhao et al. 2014; Wang et al. 2014). Different intensities of selection pressure within impaired populations may have also been responsible for varying levels of resistance with strength and breadth of resistance increasing with selection pressure. If true, this is consistent with a stronger selection pressure in the Hudson River tomcod population than experienced by the killifish populations. Each possible explanation of interpopulation differences in the degree of DE gene expression warrants further investigation.

We propose three different scenarios to explain the interpopulation differences in the number and identity of their differentially expressed genes. First, as reported previously, Hudson River tomcod exhibit a two-amino acid deletion downstream of the ligand domain of AHR2 which significantly decreases its ability to bind TCDD and reduces its effectiveness in driving reporter gene expression in AHR deficient mammalian cells treated with TCDD or PCB126 (Wirgin et al. 2011). It is also known that activation of the canonical AHR pathway directly mediates transcription of a number of xenobiotic-drug metabolizing genes in the so-called AHR battery (Nebert et al. 2000). However, the number of known genes in the AHR pathway is modest and far less than the more than one thousand that were DE between the two tomcod populations in our study. Second, it is also known that the AHR pathway interacts with a number of other downstream molecular pathways including; NF- $\kappa$ B, STAT1, liver X receptor, NRF2 (Stockinger et al. 2014; Yeager et al. 2009), estrogen receptor (Ahmed et al. 2009), hypoxia signaling (Beischlag et al. 2008), and probably others. Thus, indirect interactions of AHR-regulated proteins with components of these other pathways are additional likely sources of genes that are DE between the two tomcod populations. Third, it is also plausible that chronic exposure of the Hudson River population to mixtures of contaminants has served to reduce its responsiveness to TCDD by a mechanism independent of AHR signaling. For example, Reid et al. (2016) reported that several genes independent of AHR, and implicated in cardiovascular toxicity through disruption of voltage gated potassium channels and regulation of intracellular calcium, were selective targets of resistance to PAHs in killifish populations. However, it should be noted that disruption of

voltage-gated potassium channels had yet to be demonstrated in PCB or TCDD exposed populations. It is also possible that a more global mechanism, such as altered levels of histone modifications or DNA methylation may inhibit or activate global gene expression in highly impacted populations. We plan in future studies to investigate the possible role of epigenetic mechanisms underlying the dramatic difference in DE genes between the resistant and sensitive tomcod populations.

In summary, our characterization of the genomic sequence of Atlantic tomcod has provided a highly sensitive tool with which to begin to evaluate the magnitude of toxicities and their mechanistic bases in this impacted environmental model. A far greater number of genes than we expected were DE between the resistant Hudson River and sensitive Shinnecock Bay tomcod populations and they included many that are not normally included in the AHR battery. This study demonstrates the far greater potential extent of alteration of genomic function than expected in a vertebrate population under strong and rapid natural selection from contaminant exposure. The tomcod genomic sequence generated in this study will allow us to more fully understand the toxic impacts of exposure to the mixture of Hudson River toxicants and other stressors on this valuable environmental sentinel and to further address possible epigenetic mechanisms for these dramatic genomic differences.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

We acknowledge the Pilot Studies Program and Molecular Facilities Core of NYU National Institute of Environmental Health Sciences Center Grant ES00260, Superfund Research Program Project 5P42ES010344, Cancer Center Support Grant P30CA016087 of the Laura and Isaac Perlmutter Cancer Center at the NYUMC, NYU CTSA grant UL1TR000038 for the National Center for Advancing Translational Sciences, the Hudson River Foundation, and the Northeast Fisheries Science Center of NOAA for support of this study. We also thank members of the NYUMC Genome Technology Center for their expert assistance.

## Literature Cited

- Aken BL, et al. 2016. The Ensembl gene annotation system. Database (Oxford) pii: baw093.
- Ahmed S, Valen E, Sandelin A, Matthews J. 2009. Dioxin increases the interactions between aryl hydrocarbon receptor and estrogen receptor alpha at human promoters. *Toxicol Sci.* 111(2):254–266.
- Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

- Antkiewicz DS, Burns CG, Carney SA, Peterson RE, Heideman W. 2005. Heart malformations is an early response to TCDD in embryonic zebrafish. *Toxicol Sci.* 84(2):368–377.
- Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Beischlag TV, Morales JL, Hollingstead BD, Perdew GH. 2008. The aryl hydrocarbon receptor complex and the control of gene expression. *Crit Rev Eukaryot Gene Expr.* 18(3):207–250.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bozinovic G, Sit TL, Di Giulio R, Wills LF, Oleksiak M. 2013. Genomic and physiological responses to strong selective pressure during late organogenesis: few gene expression changes found despite striking morphological differences. *BMC Genomics* 14:779.
- Breivik K, Sweetman A, Pacyna J, Jones K. 2002. Towards a global historical emission inventory for selected PCB congeners: a mass balance approach. *Global Production and Consumption. Sci Tot Environ.* 290(1–3):181–198.
- Buchfink B, Xie C, Huson D. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12(1):59–60.
- Bugiak B, Weber LP. 2009. Hepatic and vascular mRNA expression in adult zebrafish (*Danio rerio*) following exposure to benzo-a-pyrene and 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Aquat Toxicol.* 95(4):299–306.
- Carney SA, et al. 2006. Aryl hydrocarbon receptor activation produces heart-specific transcriptional and toxic responses in developing zebrafish. *Mol Pharmacol.* 70(2):549–561.
- Courtenay S, et al. 1999. A comparison of the dose and time response of CYP1A1 mRNA induction in chemically treated Atlantic tomcod from two populations. *Aquat Toxicol.* 47(1):43–69.
- Denison MS, Nagy SR. 2003. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu Rev Pharmacol Toxicol.* 43:309–334.
- Dey WP, Peck TH, Smith CE, Kreamer G-L. 1993. Epizootology of hepatic neoplasia in Atlantic tomcod (*Microgadus tomcod*) from the Hudson River Estuary. *Can J Fish Aquat Sci.* 50(9):1897–1907.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol.* 7(10):e1002195.
- Fernandez M, Ikonomidou M, Courtenay SC, Wirgin II. 2004. Spatial variation and source prediction of PCBs and PCDD/Fs among young-of-the-year and adult tomcod (*Microgadus tomcod*) in the Hudson River Estuary. *Environ Sci Technol.* 38(4):976–983.
- Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44(D1):D279–D285.
- Gilbert D. (2013) Gene-omes built from mRNA seq not genome DNA. 7th annual arthropod genomics symposium. Notre Dame. Available from: [http://arthropods.eugenics.org/EvidentialGene/about/EvidentialGene\\_trassembly](http://arthropods.eugenics.org/EvidentialGene/about/EvidentialGene_trassembly)
- Dennis G Jr, et al. 2016. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4(5):P3.
- Hahn ME, Karchner SI, Shapiro MA, Perera SA. 1997. Molecular evolution of two vertebrate aryl hydrocarbon (dioxin) receptors (AHR1 and AHR2) and the PAS family. *Proc Natl Acad Sci U S A.* 94(25):13743–13748.
- Henny CJ, Kaiser JL, Grove RA. 2009. PCDDs, PCDFs, PCBs, OC pesticides and mercury in fish and osprey eggs from Willamette River, Oregon (1993, 2001, and 2006) with calculated biomagnification factors. *Ecotoxicology* 18(2):151–173.
- Kim D, Pertea G. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14(4):R36.
- Korf I. 2015. CEGMA. Available from <http://korflab.ucdavis.edu/datasets/cegma/#SCT5>. [2015 June 01].
- Luo R, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1(1):18.
- Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20(16):2878–2879.
- Meyer JN, DiGiulio RT. 2003. Heritable adaptation and fitness costs in killifish (*Fundulus heteroclitus*) inhabiting a polluted estuary. *Ecol Appl.* 13:490–503.
- Nebert DW, et al. 2000. Role of the aromatic hydrocarbon receptor and [Ah] gene battery in the oxidative stress response, cell cycle control, and apoptosis. *Biochem Pharmacol.* 59(1):65–85.
- Nguyen LP, Bradfield CA. 2008. The search for endogenous activators of the aryl hydrocarbon receptor. *Chem Res Toxicol.* 21(1):102–116.
- Oleksiak MF, et al. 2011. Transcriptomic assessment of resistance to effects of an aryl hydrocarbon receptor (AHR) agonist in embryos of Atlantic killifish (*Fundulus heteroclitus*) from a marine Superfund site. *BMC Genomics* 12(1):263.
- Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009. Assessing the gene space in draft genomes. *Nucleic Acids Res.* 37(1):298–297.
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14(4):417–419.
- Quadrini JD, Ku W, Connolly JP, Chiavelli DA, Israelsson PH. 2015. Fingerprinting 2, 3, 7, 8-tetrachlorodibenzodioxin contamination within the lower Passaic River. *Environ Toxicol Chem.* 34(7):1485–1498.
- Rappe C. 1993. Sources of exposure, environmental concentrations and exposure assessment of PCDDs and PCDFs. *Chemosphere* 27(1–3):211–225.
- Reid NM, et al. 2016. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* 354(6317):1305–1308.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
- Roy NK, Walker N, Chambers RC, Wirgin I. 2011. Characterization and expression of cytochrome P4501A in Atlantic sturgeon and shortnose sturgeon experimentally exposed to coplanar PCB 126 and TCDD. *Aquat Toxicol.* 104(1–2):23–31.
- Safe S, Lee S-O, Jin U-H. 2013. Role of the aryl hydrocarbon receptor in carcinogenesis and potential as a drug target. *Toxicol Sci.* 135(1):1–16.
- Seekell DA, Pace ML. 2011. Climate change drives warming in the Hudson River Estuary, New York (USA). *J Environ Monit.* 13(8):2321–2327.
- Smith CE, Peck TH, Klauda RJ, McLaren JB. 1979. Hepatomas in Atlantic tomcod *Microgadus tomcod* (Walbaum) collected in the Hudson River estuary in New York. *J Fish Dis.* 2(4):313–319.
- Stevens EA, Mezrich JD, Bradfield CA. 2009. The aryl hydrocarbon receptor: a perspective on potential roles in the immune systems. *Immunology* 127(3):299–311.
- Stockinger B, Meglio PD, Gialitakis M, Duarte JH. 2014. The aryl hydrocarbon receptor: multitasking in the immune system. *Annu Rev Immunol.* 32:403–432.
- Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, 2015. UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 3:926–932.
- Tappenden DM, Hwang HJ, Yang L, Thomas RS, LaPres JJ. 2013. The aryl-hydrocarbon receptor interaction network (AHR-PIN) as identified by tandem affinity purification (TAP) and mass spectrometry. *J Toxicol.* 213:1–13.
- Thomann RV, St. John JP. 1979. The fate of PCBs in the Hudson River ecosystem. *Ann NY Acad Sci.* 320: 610–629.

- Van den Berg M, et al. 1998. Toxic equivalency factors for PCBs, PCDDs, PCDFs for humans and wildlife. *Environ Health Perspect.* 106(12):775–792.
- Wang C, et al. 2014. The correspondence between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol.* 32(9):926–932.
- Whitehead A, Clark BW, Reid NM, Hahn ME, Nacci D. 2017. When evolution is the solution to pollution: Key principles, and lessons from rapid repeated adaptation of killifish (*Fundulus heteroclitus*) populations. *Evolution Appl.* 27:1–22.
- Wills LP, et al. 2010. Comparative chronic liver toxicity of benzo[a]pyrene in two populations of the Atlantic killifish (*Fundulus heteroclitus*) with different exposure histories. *Environ Health Perspect.* 118(10):1376–1381.
- Wirgin I, Chambers RC, 2006. Atlantic tomcod (*Microgadus tomcod*): a model species for the response of Hudson River fish to toxicants. In: Waldman JR, Limburg KE, Strayer D, editors. *Hudson river fishes and their environment.* Am Fish Soc Symp 5:331–364.
- Wirgin I, Waldman JR. 2004. Resistance to contaminants in North American fish populations. *Mutat Res.* 552(1–2):73–100.
- Wirgin I, et al. 1992. Effects of prior exposure history on cytochrome P450A mRNA induction by PCB congener 77 in Atlantic tomcod. *Mar Environ Res.* 34(1–4):103–108.
- Wirgin I, Weis JS, McElroy AE, 2006. Physiological and genetic aspects of toxicity in Hudson River species. In: Levinton J, Waldman JR, editors. *The Hudson river ecosystem.* Cambridge: Cambridge University Press. p. 441–464.
- Wirgin I, et al. 2011. Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson River, USA. *Science* 331:1322–1325.
- Xie Y, et al. 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30(12):1660–1666.
- Yasui T, et al. 2007. Functional characterization and evolutionary history of two aryl hydrocarbon receptor isoforms (AhR1 and AhR2) from avian species. *Toxicol Sci.* 99(1):101–117.
- Yeager RL, Reisman SA, Aleksunes LM, Klaassen CD. 2009. Introducing the “TCDD-inducible AhR-Nrf2 gene battery. *Toxicol Sci.* 111(2):238–246.
- Yuan Z, Courtenay SC, Chambers RC, Wirgin I. 2006a. Evidence of spatially extensive resistance to PCBs in an anadromous fish of the Hudson River. *Environ. Health Perspect.* 114:77–83.
- Yuan Z, Courtenay S, Wirgin I. 2006b. Comparison of hepatic and extra hepatic induction of cytochrome P450A by graded doses of aryl hydrocarbon receptor agonists in Atlantic tomcod from two populations. *Aquat Toxicol.* 76:306–320.
- Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9(1):e78644.

Associate editor: Mandë Holford